

XamForumDB: a dataset for studying Q&A about cross-platform mobile applications development.

MATIAS MARTINEZ, SYLVAIN LECOMTE
UNIVERSITY OF VALENCIENNES, FRANCE
FIRST NAME.LAST NAME@UNIV-VALENCIENNES.FR

Android and iOS are the two mobile platforms present in almost all smartphones build during last years. Developing an application that targets both platforms is a challenge. A traditional way is to build two different apps, one in Java for Android, the other in Objective-C for iOS. Xamarin is a framework for developing Android and iOS apps which allows developers to share most of the application code across multiple implementations of the app, each for a specific platform. In this paper, we present XamForumDB, a database that stores discussions, questions and answers extracted from the Xamarin forum. We envision research community could use it for studying, for instance, the problematic of developing such kind of applications.

ACM Reference format:

Matias Martinez, Sylvain Lecomte
University of Valenciennes, France
first name.last name@univ-valenciennes.fr . 2016. XamForumDB: a dataset for studying Q&A about cross-platform mobile applications development.. 1, 1, Article 1 (January 2016), 8 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Nowadays, there are billions of smartphone devices around the world which run Android or iOS operating systems. A *cross-platform mobile application* is an application that targets more than one mobile platform. For example, most of the famous social apps are available for both Android and iOS platforms.

A traditional approach for developing this kind of apps is to build a *native* application for each platform. Here, each native apps is build using a specific programming language (e.g., Java for Android, Objective-C vs Swift for iOS), SDK and IDE (e.g., Android Studio, XCode for iOS). Unfortunately, this development process increases costs of development and maintenance phases. For example, the mobile development team must be competent to each target mobile platform.

During last years, researches and industry sectors have both focused on proposing solutions for developing mobile applications with the goal of overcoming the mentioned problematic. One of those solutions is Xamarin¹, a framework to build native Android, iOS and Windows development in C# programming language.

¹<https://www.xamarin.com/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2016 ACM. Manuscript submitted to ACM

² The advantage of Xamarin platforms is that developers re-use their existing C# code, and share significant code across device platforms. For example, the app Evolve² for managing conferences' events, was developed using Xamarin and is around 15,000 lines of code: the iOS version contains 93% shared code i.e., C# (the remain 7% iOS specific code) and the Android version contains 90% shared code³.

Xamarin has an official online forum⁴, which has multiple purpose: a) to be a platform for communicating announcements (events, releases, jobs related with the platform) and b) to be a Q&A site, where Xamarin users (such as cross-platforms developers) post discussions, questions and share answers. Also, the developers of the Xamarin platforms (i.e., members of the company that develop it) participate on the forum.

However, to the best of our knowledge, no work studies the quality of cross-platforms mobile apps, nor analyzes this source of information for understanding the development process of mobile apps using a cross-platform development tool such as Xamarin.

In this paper, we present XamForumDB, a database that contains all discussions, questions and answers from the Xamarin forum. Moreover, it stores the number of views of each post, number of 'likes', the accepted answers, and profile of the forum's user.

We consider that XamForumDB gives to the software engineering research community the possibility to study the development and maintenance phases of cross-platform mobile applications, and to propose new approaches for improving the quality of mobile applications. The XamForumDB is publicly available at <https://goo.gl/A1eh03>

2 DESCRIBING FORUM XAMARIN

The Xamarin forums site is a online web platform where mobile developers post questions or start a discussion about the Xamarin framework and its ecosystem. Moreover, it is a communication channel between the developers of the framework (a.k.a. Xamarin Team) and users of it (i.e., the developers of mobile apps that use Xamarin as development framework). For instance, new version of the framework or particular components are announced in the forum.⁵ All posts are publicly available. However, for creating a new post, a user has to register in the site, which is free.

2.1 Forums

The Xamarin Forums platform is composed of seven general forums: Community, General, Pre-release & Betas, Tools and Libraries, Graphics & Games, Xamarin Platform and Xamarin Products. For instance, the forum Xamarin Platform contains discussion and questions related to the development of an application for different targeted applications, while the forum Xamarin Product focuses on discussions about products related to the Xamarin technology such as TestCloud, a platform for testing the mobile app built using Xamarin.

Each forum has one or more 'sub-forums'. For instance, the mentioned Platform forum has 5 categories: Android, iOS, Cross-Platform, Mac, and Xamarin.Forms. The forums are established in advance by the forum administrators, which means that users are not able to define new ones.

We identify two types of forums: some related to technological topics (platforms, libraries, tools, IDEs) and others related to non-technological topics related to Xamarin (focus on event, conferences, jobs, etc.)

²<https://github.com/xamarinhq/app-evolve>

³<https://goo.gl/eqj1ml>

⁴<https://forums.xamarin.com/>

⁵<https://forums.xamarin.com/discussion/85747/xamarin-forms-feature-roadmap#latest>

The main page of each forum shows a paged-list of threads (i.e., posts) and two buttons, one for creating a question and the other for creating a discussion. Each thread from the list shows the post's title, author, number of views, number of answers, and zero or more labels. Those labels are colored boxes located near the title and indicate, for instance, if a question was answered, if the answered was accepted by the user who wrote the question, or if a post is an announcement.

2.2 Posts

A forum user can create a post for a given category. Even there is two types of post, i.e., questions and discussions, the creation forms of them are similar, i.e., they have the same fields. The user can also select a list of tags associated his post. The forum suggests popular tags such as “Android”, “async” or “httpClient”. However, once a post is created, the forum does not show the list of tags associated to each post.⁶

2.3 Answers

A registered user can answer an existing question or to put a ‘like’ on existing answers. Moreover, as in others Q&A such as Stack Overflow, the post's author can *accept* one or more answers. The accepted answers are labeled with a green-colored box.

2.4 Users

The forum has the functionality for registering new users, which have the right to create new posts and to write comments. However, neither questions nor answers can be erased by a registered user once they are submitted. In addition to a registered user, some users correspond to the Xamarin Team, which belong to the company that develops Xamarin framework, acquired by Microsoft in 2016.

3 DATASET CONSTRUCTION

The Xamarin forum does not have an API to programmatically access and retrieve the forum data, such as that one provided by Stackoverflow⁷ or GitHub⁸. For this reason, we developed a web crawler for retrieving and storing all data the Xamarin forum presents. Our engine has main two phases: Page fetching and Page parsing.

3.1 Fetching pages

The first phase fetches (i.e., downloads) web pages written in HTML for the forum site. The engine accesses to those pages via HTTP protocol. We fetched two kinds of pages: The first kind are pages that correspond to a single post. This pages contain post's title, question (or discussion topic), author data (names, location, roles), posting date and a list of comments. Each comment has the user name that wrote it, the date, and one or more labels such as “Answered question”. The second kind of pages corresponds to the main pages of each forum. A “main” page shows in a paged list all posts done in that forum.⁹ The list shows first the post corresponding to announcements, then the remaining posts ordered by decreasing creation date. Moreover, the list shows for each post its title, the numbers of views (i.e.,

⁶Last visit: February, 2017

⁷<https://api.stackexchange.com/>

⁸<https://developer.github.com/v3/>

⁹In those main pages, those post are called “Threads”

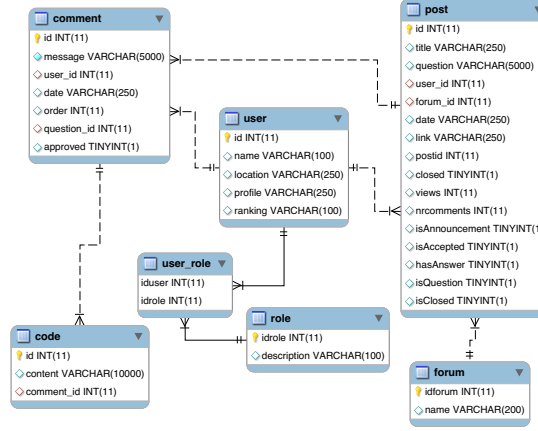


Fig. 1. Entity-relationship model of XamForum.

number of visits the post has received), numbers of comments done, and zero or more labels that indicates if the post is a question or announcement, if it has an accepted answer, etc.

3.2 Extracting data from pages

Then, in the second phase, our engine extracts data for each fetched page. It parses, search and reformat data from the two kind of pages fetched. The data we extracted is that one mentioned in the previous phase. Then, the extracted data is stored in a relational database.

Our engine is implemented in python language, and uses the library BeautifulSoup¹⁰ for parsing the Xamarin web pages in HTML. The structured data is then stored in a MySQL database using the ORM framework SQLAlchemy¹¹ for connecting our Python engine app with the database.

3.3 Data Schema

Figure 1 shows the schema of our database which stores all data we extracted from the Xamarin forum. In this section, we explain the main entities from that schema.

The first entity is **Post**, which stores all posts done in all Xamarin forums (right in the figure). Each post belongs to one **Forum** and has zero or more **Comments**. Note that the post entity stores both questions and discussions. We decide to model both questions and discussion in the same entity due they share the majority kind of data. The kind of post can be inferred from the presence or absence of labels, which are modeled as boolean fields in table **Post**. For instance, the field ‘isAccepted’ indicates if a post has at least one comment accepted. We decide to delegate the post classification task to the database users.

Both **Comment** and **Post** entities have a relation *many-to-one* with the entity **User**, which stores the information of the users that make or answer questions. A **User** has different **Roles**. Registered user has the role ‘Member’ by default, but there are other roles exclusively assigned to used belonging the Xamarin organization such as ‘Forum

¹⁰<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

¹¹<http://www.sqlalchemy.org/>

Administrator’, or ‘XamUProfessor” which are professionals involved on the Xamarin training program called Xamarin University.¹²

Finally, the entity **Code** corresponds to code chunks the users write on comments. One comment could have one or more code chunks. We remark that this entity do not only represent programming code chunks (e.g., C# or Java code) but also execution traces. We let users of XamForumDB to classify the data stored in **Code** entity.

4 DATA ANALYSIS

In this section, we briefly mention the amount of the data that XamForumDB stores.

4.1 Posts and Users

The database has 75,380 posts. The first dates from January, 29 2013, and the last one from January, 17 2017, date we finished to carry out the data extraction from the Xamarin forum. The number of Users that have done at least one question or one post is 32,798. The large majority (31,310) are users registered in the forum site, i.e., they are external to the Xamarin organization. Between the remaining users there are, for instance, 164 users from the Xamarin development team, and 41 forum administrators.

4.2 Comments

The database XamForumDB stores 54,542 posts with at least one comment, while 20,838 posts have not received any comment, and around 15,000 posts (around 20% of all posts) has only one comment. The maximum number of comments done in a single post is 41. In total, the XamForumDB stores 226,927 comments.

4.3 Forums and Posts

Table 1 shows all Xamarin forums with the number of posts (questions or discussion) that each one has. We observe that the majority of posts belong to “technological” forums: almost the 30% are related to the Xamarin.Forms technology (a Xamarin API for developing simple UI in C# or XAML languages), and the 23% and 19.5% of all posts belong to Android and iOS forums, respectively. The fourth forum with most number of posts is about Xamarin Studio, an IDE for developing Xamarin application based on the Microsoft Visual Studio. Between the non-technological forums, Events has 954 posts (1.27%), where Job Listings 376 (0.5%).

4.4 Code Chunks

The number of total code chunks in the Xamarin forum is 73,304. Those chunks appear in 37,321 comments, that means the 16% of comments have at least one code chunk, whereas 19,673 posts have at least one chunk in any comment (representing the 36% out of the commented posts).

5 RESEARCH OPPORTUNITIES

In this section we present some research directions that, in our opinion, could be done by analyzing the data from XamForumDB.

First, by analyzing discussions and questions from XamForumDB, researchers could understand the main concerns of developing cross-platform mobile applications using a state-of-the-art development framework such as Xamarin. For instance, researchers could mine the most frequent problematic or bugs that developers face during development,

¹²<https://www.xamarin.com/university>

Forum name	#Posts	%
Xamarin Platform-Xamarin.Forms	22478	29.82%
Xamarin Platform-Android	17635	23.39%
Xamarin Platform-iOS	14702	19.50%
Tools and Libraries-Xamarin Studio	3999	5.31%
Xamarin Platform-Cross Platform	3370	4.47%
General	3148	4.18%
Tools and Libraries-Visual Studio	2777	3.68%
Xamarin Platform-Mac	1708	2.27%
Community-Events	954	1.27%
Xamarin Products-Xamarin Test Cloud	880	1.17%
Tools and Libs-Components, and Plugins	849	1.13%
Graphics & Games-CocosSharp	636	0.84%
Xamarin Products-Xamarin Insights	601	0.80%
Prerelease & Betas	443	0.59%
Community-Job Listings	376	0.50%
Community	175	0.23%
Tools and Libraries-Profiler	154	0.20%
Graphics & Games-UrhoSharp	152	0.20%
Tools and Libraries-Objective Sharpie	123	0.16%
Tools and Libraries-Workbooks & Inspector	82	0.11%
Community-Presentations	64	0.08%
Xamarin Platform-Xamarin.Forms Evolution	53	0.07%
Others forums	21	<0.03%
Total	75380	

Table 1. The table presents all Xamarin forums and the number of posts each forum includes.

and propose frequent bugs fix patterns mined from answers, questions and the code attached to them. Moreover, as the code chunks from XamForumDBinclude execution traces, it would allow researchers to analyze commons failure during the execution of cross-applications.

Researchers could also compare efforts of developing iOS apps against developing Android apps using the Xamarin framework.

The XamForumDB includes forums that discuss about tools for developing cross-platform mobile applications such as Xamarin Studio. Then, researches could analyze the *usability* of those tools from the comments written in those forums, for understanding their limitations and even propose new ones or extensions.

Even Xamarin forum is the official forum, hosted and administrated by that organization, it's not the only Q&A site where developers post question about Xamarin platforms. Researcher could mine similar questions posted in two or more Q&A sites, (e.g., Xamarin forum and Stack Overflow) and compare the answers found in those sites to, for instance, cross-validate accepted answers.

6 LIMITATIONS

Our fetching engine is not capable of capturing all information the Xamarin forum presents to a registered user. For instance, a registered user can “like” an existing answer. The Xamarin forum shows the number of like for each answer *only* to registered user. As our fetching engine inspects the pages to fetch as an “anonymous user” (i.e. it is not logged to the forum site) then the number of likes is not presented. Moreover, registered user can post images in the comments. The current version of our engine is not capable of retrieving and storing them in our database.

7 RELATED WORKS

The Mining Software Repository (MSR) community has released a dump of the Q&A site Stack Overflow [1] for the mining challenge competition of 2013. Several works have been proposed by analyzing that dump. For instance, Baselli et al.[3] studied the personal trails of stack overflow users. Research have been focused on finding frequent questions for targeting a particular problematic. For example, Pinto et al. have mined questions to first characterize energy consumption problems [8].

Several works have applied topic modeling technique such as latent Dirichlet allocation (LDA). For example, Baura et al. [2] analyzed Stack Overflow data to automatically discover the main topics present in developer discussions and analyze the topic popularity over the time.

Other works have focused on the conjunction of mining Q&A and mobile technologies topics. For instance, Linares-Vasquez et al. also used LDA to extract hot-topics from mobile-development related questions [7]. Their findings suggest that most of the questions include topics related to general questions and compatibility issues. Moreover, Linares-Vasquez et al.[6] studied questions and activity in Stack Overflow when Android APIs occurs. They found that deleting public methods from APIs is a trigger for questions that are discussed more and of major interest for the community. Rosen et al. [9] applied (LDA) based topic models on the mobile-related Stack Overflow posts to determine what mobile developers are asking. Across all platforms studied, they found that questions related to app distribution,user interface, and input are among the most popular.

Beyer et Pinzge [4] investigated 450 Android related Stack Overflow posts to get insights into the issues of Android app development. They found that the problems that are discussed most often are related to ‘UserInterface’ and ‘Core Elements’. Moreover, the authors presented approach to group tag synonyms to meaningful topics[5]. Using their approach, they reduced the number of 38,000 diverse tags on Stack Overflow to 2,481 meaningful tag.

Other works focus on analyzing developer forums. For example, Venkatesh et al. [10] mined both developer forums and Stack Overflow to find the common challenges encountered by client developers that use Web APIs.

8 CONCLUSION

We present XamForumDB, a database that contains structured data such as discussions, questions, answers, user and their roles, extracted from the Xamarin forum, a online site where developers that use the Xamarin framework for building their cross-platform mobile apps interact. As developing such kind of applications is a challenge due to divergences of the mobile platforms, we envision the research community use XamForumDB to better understand how is the development and maintenance of cross-platform mobile applications. The XamForumDB is publicly available at <https://goo.gl/A1eh03>

REFERENCES

- [1] Alberto Bacchelli. Mining challenge 2013: Stack overflow. In *The 10th Working Conference on Mining Software Repositories*, page to appear, 2013.
- [2] Anton Barua, Stephen W. Thomas, and Ahmed E. Hassan. What are developers talking about? an analysis of topics and trends in stack overflow. *Empirical Softw. Engg.*, 19(3):619–654, June 2014.
- [3] Blerina Bazelli, Abram Hindle, and Eleni Stroulia. On the personality traits of stackoverflow users. In *Proceedings of the 2013 IEEE International Conference on Software Maintenance, ICSM '13*, pages 460–463, Washington, DC, USA, 2013. IEEE Computer Society.
- [4] S. Beyer and M. Pinzger. A manual categorization of android app development issues on stack overflow. In *2014 IEEE International Conference on Software Maintenance and Evolution*, pages 531–535, Sept 2014.
- [5] Stefanie Beyer and Martin Pinzger. Grouping android tag synonyms on stack overflow. In *Proceedings of the 13th International Conference on Mining Software Repositories, MSR '16*, pages 430–440, New York, NY, USA, 2016. ACM.

- [6] Mario Linares-Vásquez, Gabriele Bavota, Massimiliano Di Penta, Rocco Oliveto, and Denys Poshyvanyk. How do api changes trigger stack overflow discussions? a study on the android sdk. In *Proceedings of the 22Nd International Conference on Program Comprehension, ICPC 2014*, pages 83–94, New York, NY, USA, 2014. ACM.
- [7] Mario Linares-Vásquez, Bogdan Dit, and Denys Poshyvanyk. An exploratory analysis of mobile development issues using stack overflow. In *Proceedings of the 10th Working Conference on Mining Software Repositories, MSR '13*, pages 93–96, Piscataway, NJ, USA, 2013. IEEE Press.
- [8] Gustavo Pinto, Fernando Castor, and Yu David Liu. Mining questions about software energy consumption. In *Proceedings of the 11th Working Conference on Mining Software Repositories, MSR 2014*, pages 22–31, New York, NY, USA, 2014. ACM.
- [9] Christoffer Rosen and Emad Shihab. What are mobile developers asking about? a large scale study using stack overflow. *Empirical Softw. Engg.*, 21(3):1192–1223, June 2016.
- [10] P. K. Venkatesh, S. Wang, F. Zhang, Y. Zou, and A. E. Hassan. What do client developers concern when using web apis? an empirical study on developer forums and stack overflow. In *2016 IEEE International Conference on Web Services (ICWS)*, pages 131–138, June 2016.